A Visual Representation Model for FMRI Activity Decoding

Piyawat Saengpetch[1]
and Luepol Pipanmemekaporn[2]
Department of Computer and Information Science,
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
[1]piyawat@sru.ac.th
[2]luepol.p@sci.kmutnb.ac.th

Received: 24/9/2019
Accepted: 16/10/2019

Abstract - **This study presents how visual features can be utilized for predicting arbitrary concepts from brain activity patterns. Six types of low-level image features, including: (1) color histogram, (2) color correlogram, (3) edge direction histogram, (4) color moments, (5) wavelet texture, and (6) Scale Invariance Feature Transform are extracted from online images associated with these concepts where (1)-(3) are based on global features and (4)-(6) are based on local ones. We also present a method of selecting images associated with each concept from Flickr, a large social image resource, based on tag relationship. Experimental results conducted on fMRI dataset provided by Carnegie Mellon University demonstrated that local-based visual feature models achieve encouraging performance of the task of predicting arbitrary words from fMRI activity patterns compared to global-based feature models and have no significant difference with state-of-the-art text-based feature models with manual user.**

Keywords - **Brain Decoding, Visual Features, FMRI, Concept Prediction and Vector Space Model**

I. INTRODUCTION

Neuroimaging is applied in the medical domain to diagnose neurological and cerebrovascular disease. In recent studies, a variety of approaches have been developed for neurological diagnosis, including Magnetoencephalography (MEG) and Computed tomography (CT). Among these techniques, Functional Magnetic Resonance Imaging (fMRI) measures brain activity patterns via detecting oxygen in blood cells (Blood Oxygen Level Dependent: BOLD). During the brain is activated, blood cells level is increased [2]. Then, brain activity patterns were converted into a 3D image beneficial in Neuroscience research, Artificial Intelligence, and Brain-Computer Interface systems [1]. Brain activity patterns such as reading, thinking, emotions, and so on that appear approximate 20,000 voxels in the fMRI image.

According to Mitchell et al. [3] presented the predicting brain stimulation patterns technique for fMRI associated with each concept. This approach investigated a computational model that describes the meaning of the relationship between brain voxels and nouns through machine learning method. Thus, brain voxels play an important role in fMRI analysis [5-6]. Furthermore, Palatucci et al. [4] presented a model for predicting the concept of brain activity patterns. The performance of the models depends on the representation concept scheme or concept space. The concept space is generated based on semantic relationship between nouns and 25 verbs by linguists in [3], 218 questions have been studies to defined the conceptual space [4]. Although these models give relatively high accuracy, the limitations of these models do not support a variety of concepts. To solve this problem, several existing literature used Natural language processing (NLP) to extract text or word features from large text libraries such as Wikipedia [6], Google n-gram corpus [4] and 50 million web pages [5]. However, the efficiency of these methods is still not very high. Therefore, the problem is still challenging in this application domain.

To overcome this shortcoming, in this paper, the analysis of visual features is applied instance NLP approach. The advantages of visual features are:

- Easy low-level image processing.
- Recent research shown benefits of using peculiar characteristics to solve NLP problems [7-8].
- It can be described concepts in a view that cannot be explained by language such as colour and texture [8].

This paper extracted 6 types of visual features that were used for describing concept of images as below; 1) color feature such as color histogram, auto-correlogram, and color moments and 2) texture characteristics such as edge direction histogram, wavelet texture, and Scale Invariance Feature Transform (SIFT). These features are used to describe concepts in terms of lexical in order to support a variety of concepts. We selected related images from online gallery Flickr. Flickr currently has more than 14 million images with tags that is identified by internet users. These are characterized by visual features from the fMRI data of the Carnegie Mellon University [3]. The results show that the proposed visual features are efficiently compared to text features.

II. REPRESENTATION MODEL FOR BRAIN DECODING

The most regularly used models for brain decoding were done by using semantic knowledge to improve the class label. In other words, it utilizes an intermediate set of features obtained from the semantic knowledge for characterizing a large number of classes. Figure 1 shows the typical brain decoding model [4]. The model consists of 3 parts: 1) input layer instead of voxel activation vector $X = \{x_1, x_2, \ldots, x_n\}$, 2) intermediate layer is represented by the characteristic vector to describe the concept. $Z = \{z_1, z_2, \ldots, z_k\}$, and 3) decoding layer represented by the target concept vector $W = \{w_1, w_2, \ldots, w_m\}$. The relationship between the middle layer and decoder layer is explained by the characteristic value which is associated with each concept. For the relationship between stimulus vector of brain points and features can be described as the regression model (1).

$$\vec{W} = (X^T X + \lambda I)^{-1} X^T Z \tag{1}$$

Let $\vec{W}, I$ be a coefficient matrix and identity matrix. λ be regularization parameter where λ is obtained by cross validation.

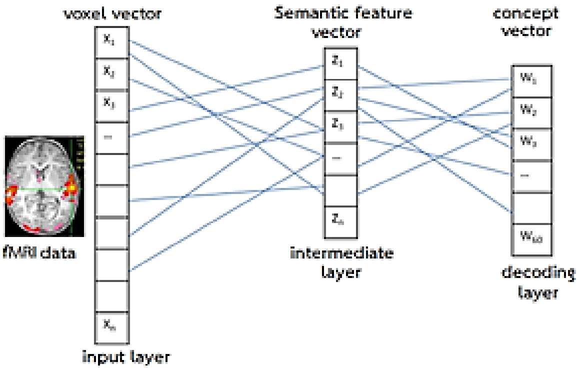


Figure 1. Generic Model of Brain Decoding

In order to predict the concept. We first examined fMRI image to represent the activated brain vector X. This vector is used to visual feature *Zˆ(X)*. Then, visual feature is measured cosine similarity in each concept *Z(w')* and an answer is selected from the most similarity concept.

$$\underset{w' \in W}{argmax}\ cos\left(\hat{Z}(X), Z(w')\right) \tag{2}$$

For the vector Z in each concept, the method [4] proposed describing concept technique using 218-dimensional which are the questions of the concept characteristic such as “Can you hold it?”, “Is it animal?”, and many more. Although, these features only predict specific designed concept group, but cannot apply in increasing concept. Moreover, many literatures conducted the automatic visual feature extraction approach for a concept describing using linguistic knowledge and text corpus. For example, the method [4] used a semantic associated with each concept from Wikipedia to represent vector *Z*. The paper [5] has defined that vector *Z* is a co-occurrence vector between any concepts that can be found in English corpus 50 million pages. However, the results shown that the prediction accuracy is not very high compared to state-of-the-art text-based feature model with a manual user.

III. VISUAL SEMANTIC MODELS

In this paper, we analysed low-level features to predict the conceptual. Low-level features will be described the relationship between concept and NLP such as color and texture. Moreover, previous works shown low-level image features capability in lexical concept [7-8].

A. Colour-Based Features

Colour is an important visual perception and interpretation. Visual color feature is the fundamental which is used in image describing. For this paper, visual color feature consists of following.

- **Color histogram** [11] is carried out by color distribution which is related with frequency due to its various color systems. We employ a LAB color histogram. Let L is brightness, A is ratio of red and green, and B is ratio of blue and yellow. Due to resistance to noise and to reduce dimensional of the color histogram, we devised an equivalent range in each LAB into four levels and count frequency in each level. Thus, in this work, 64 color histograms have been used.

- **Color correlogram** Let p_i and p_j is the pixel in images that distance $|p_i-p_j|=k$ spatial color relation C_x and C_y is defined the probability of color C_x in pixel p_i distance k from pixel p_j is C_y. In this paper, we used color correlogram HSV which is represented by [12] range of color 36 levels and distance set $k=\{1, 3, 5, 7\}$. Thus, visual feature based on color correlogram is 36x4 = 144-dimensional.

- **Color moments** [11] used to represent statistical value: mean, variance and skewness in each color component 3x3x3 vector (9 dimensional). However, it is not well for image classification. In this paper, we generated a grid 5x5 pixels across image and compute color moments in each grid and compute color moment in each grid. As a result, this technique is obtained 5x5x9 = 225 dimensional.

B. Texture-Based Features

In digital image processing, texture is defined as a foundation that appears in image. Texture can be represented the image content by examining substructure, which is resistant to noise and provides image detail compared to color feature. In this article, texture feature is examined by following;

- **Edge direction histogram** [13] represent distribution of edge direction in images. In general, edge direction histograms have 73-dimensional. In first 72-dimensional represent edge direction which is divided in each 5 degree. The remaining dimensional represent amount of pixel which is not detected. In this work, we proposed Canny edge detector and Sobel operator to find edge direction by gradient computation in each pixel.

- **Wavelet texture** [14] is a popular color transforming technique that is discrete 2-dimensional by decomposition of images into multilevel sub-band. In each level, the original image is divided into 4 bands (LL, LH, HL, and HH). "L" is low frequency and "H" is high frequency. Two transforming types are Pyramid-structured Wavelet Transform (PWT) for LL decomposition and Tree-structured wavelet transform (TWT) is used to LH, HL and HH decompose at each level. Then, the visual vector is computed from the mean and standard deviation of the wavelet coefficient which is obtained multilevel sub-band. In this paper, we assigned 3 deep levels. Thus, PWT is 24 (3x4x2) visual features and TWT is 104 (52x2) visual features and total is 128-dimensional.

- **Bag-of-words of Scale Invariance Feature Transform (SIFT**) [15] is visual feature for image analysis. SIFT is widely used in many approaches which is feature invariant to image translation, orientation, scaling, and partially in variant to illumination changes. SIFT features are efficiency two main stages: scale-space extrema detection and keypoint localization. Key location in scale space is implemented by looking for difference of Gaussian function. At each candidate location, keypoints are selected to describe image. In general, keypoints are represented 128-dimensional. However, there are similarity various of keypoints. The reason for this, *k*-mean technique has been used to divide group of keypoints and then representation into k dimensional and so called bag-of-words. In this paper, SIFT features are 500-dimensional.

C. Image-Based Semantic Vectors

In this paper, we used the NUS-WIDE [9] image to extract visual features for concept describing. NUS-WIDE dataset has 269,648 images that are collected from Flickr. NUS-WIDE is related to the tag which is used to describe the details in image. The total is 5,018 tags that are used more than 100 times. The average is 8.5 tags. Let $W=\{w_1,w_2,...,w_m\}$ is set of m concept. We analyzed images set from NUS-WIDE which is tagged over 4 words and at least 1 tag as a concept. However, there are many images to extract features and each concept describe in various images. Therefore, we present image selection technique based on the relationship between tag and concept. In such a technique, the concept with high frequency tag is selected in this work. Let $T=\{t_1,t_2,...,t_n\}$ is the most set of tag frequency n of NUS-WIDE ($n=1,000$). Tag relationship t_i concept w_j defined as equation (3).

$$r_{ij} = \frac{f_{ij}}{f_i+f_j-f_{ij}} \quad (3)$$

where f_{ij} is the tag frequency t_i and concept w_j appears in the same image where the value r_{ij} will be between 0 and 1. After the relationship between tag and concept is obtained, each image *I* will be used to select images for each concept using the equation (4).

$$S(w_j, I) = \sum_{t_i \in T(I)} r_{ij} \quad (4)$$

where $S(w_j,I)$ is score of image I , concept w_j $T(I)$ is set of tag 1k otherwise I. Maximum score k in concept is chosen to extract visual feature k. We assigned k=100 in the propose. Visual feature in each concept is computed mean and normalization by using z-score, then brain encoding model is generated as detailed in section 2.

IV. EXPERIMENTS

In each section, we detailed the dataset, experimental and results.

A. fMRI Dataset

We utilized the fMRI dataset from Carnegie Mellon University. The participants are 9 right-handed adults (P1-P9) which were stimulated through drawings and then labeled into 60 concrete objects. These concepts were presented 6 times 6x60=360 and then 60 classes average in each concept. The fMRI images are transformed into activate vector of the voxel that is 20,000 voxels on average. In the method [3] and [4] fMRI vectors are reduced to 500-dimensional. In experimental, we used such a technique in the model. Thus, the searchlight algorithm is used.

B. Baselines

In our research, we compared the performance of semantic vector which is obtained from visual feature from the document and text-based feature.

- **Verb25** is presented in [3] which is described the concept in form of the noun and describing the co-occurrence vector between nouns and verbs of 25 words such as “see”, “hear”, “taste”, “smell”, “eat” and so on. These general verbs are often concrete nouns in English sentences and designed by linguists. In the experiment, Verb25 is an effective result in this dataset.

- **Human218** is presented in [4] which is described as the concept in the form of 218-dimensional. The vector is obtained from 218 questions for example, “*Can you hold it?*”, “*It is a manmade?*” or “*It is animal?*” etc. These questions were designed by linguists and collected answers based on cloud sourcing and then compute mean answers which are related to the concept.

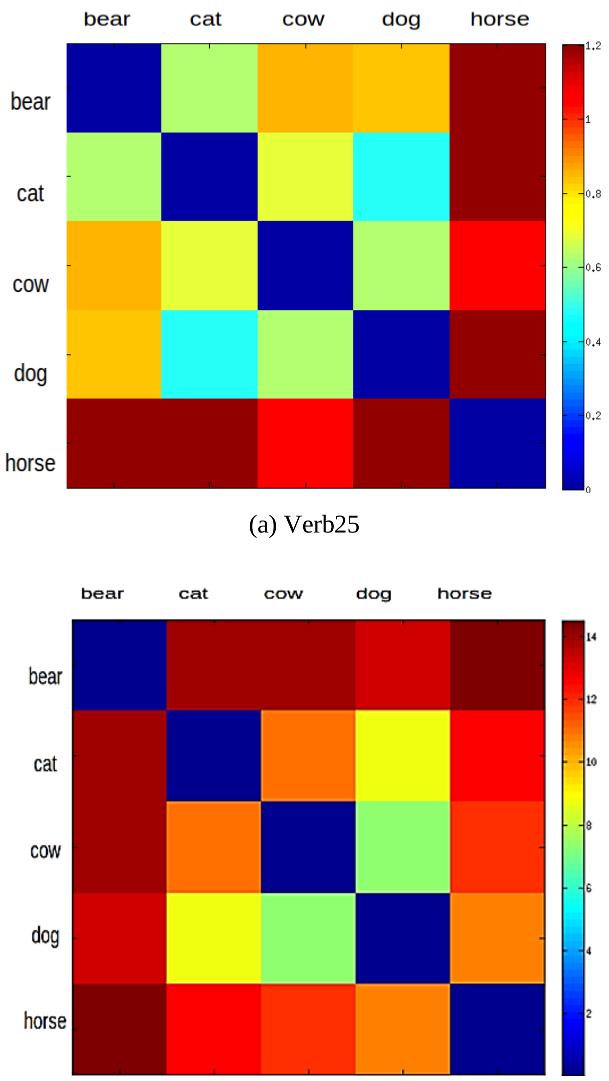
C. Experimental Setting

From above mentioned, various concept describing is an essential for visual semantic vector. Research [5] presented Leave-two-out cross-validation approach by training model from 58 fMRI images (1 image: 1 concept). Then, 2 fMRI images are used to test the accuracy. This result are compared to visual feature in each concept. The answer is obtained based on similarity measurement. The procedure is interactively training $\binom{60}{2} = 1,770$ times. 60 concepts are compared to the result 1,770 times x 2 concepts = 3,540 times. Therefore, the concept prediction accuracy is defined the ratio between corrected prediction and total prediction. In this experimental, we used leave-two-out cross-validation in visual feature comparison.

D. Results and Discussion

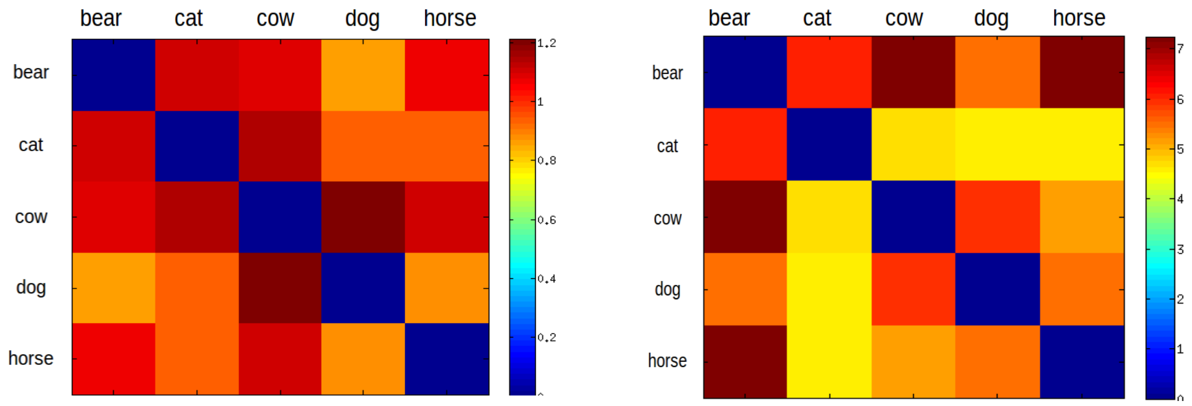
In this paper, six vector visual features: color histogram (CH), color correlogram (CORR), color moment (CM), wavelet texture (WT), edge direction histogram (EDH) and Bag-of-word of SIFT (BoW) are compared to Verb25 and Human218, A feature vector derived from the text. The comparison of the model efficiency based on various methods as shown in Table I, it shown that BoW, which belongs to the surface-related visual model, provides the predictive accuracy of 60 concepts for the P1 to P9. Following by Human218 and Verb25, which are generated based on linguistic features influenced by BoW [3-5]. As a result, visual texture features is powerful prediction more than color feature except WT provides the lowest accuracy.

To order to present semantic vector for describing the similarity conceptual. We selected BoW and EDM compared to Verb25 and Human218 to distinguish between the concepts in the animal group consisting of "bear", "Cat", "cow", "dog" and "horse", based on Cosine similarity measurement (Figure 2). BoW and Human218 can describe the concept with no significance.



(a) Verb25

(b) Human218



(c) EDM

(d) BoW

Figure 2. Correlation among Concepts in Animal Groups

TABLE I
COMPARISON OF THE ACCURACY (%) OF THE MODELS CREATED BY DIFFERENT FEATURES VECTOR.

Category	Method	P1	P2	P3	P4	P5	P6	P7	P8	P9	Mean Rank	Rank
Color-based feature	CH	44.69	42.03	41.64	38.22	38.19	28.31	37.32	30.71	43.14	4.78	5
	CORR	42.60	41.05	36.58	36.05	37.68	26.30	34.58	34.04	39.01	5.89	6
	CM	27.26	23.50	24.89	29.15	21.69	17.99	25.54	18.42	23.84	7.00	7
Texture-based feature	WT	20.37	17.74	17.71	16.84	16.75	15.56	16.53	15.88	17.37	8.00	8
	EDH	43.87	47.15	40.45	45.65	42.06	29.41	36.38	36.13	45.11	4.11	4
	BoW	**57.26**	49.58	**54.60**	53.67	52.12	49.15	44.80	**60.59**	**51.67**	1.89	**1**
Text-based features	Verb25	56.30	53.19	53.14	**54.58**	**53.28**	42.94	47.97	46.05	45.00	2.22	**3**
	Human218	50.40	**55.37**	46.69	43.98	52.40	**49.41**	**60.96**	46.53	46.53	2.11	**2**

V. CONCLUSIONS

This paper, we proposed a new method for concept prediction based visual features through image processing from fMRI. Our approach selected 6 visual features such as: (1) color histogram, (2) color correlogram, (3) color moments, (4) edge direction histogram, (5) wavelet texture, and (6) Bag-of-Words (BoW) of SIFT, where (1)-(3) is a distinctive color feature, while (4)-(5) is a distinctive image surface feature. We designed the experiment and compared our results with human218 and Verb25. As reported in the experimental results, the BoW technique has the best results.

VI. ACKNOWLEDGEMENTS

This research was funded by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. Contact no. 6346102.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Sitaram, R., Weiskopf, N., Caria, R., Veit, M., & Birbaumer, N. (2008). fMRI brain-computer Interfaces. IEEE Signal Process. Mag., 25(1), 95-106.

[2] Norman, K.A. & et al. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in cognitive sciences, 10(9), 424-430.

[3] Mitchell, T., Shinkareva, S.V., Carlson, A., Chang, K., Malave, V.L., Mason, R.A., & Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. Science, 320(5880), 1191-5.

[4] Palatucci, M., Hinton, G.E., Pomerleau, D., & Mitchell, T.M. (2009). Zero-Shot Learning with Semantic Output Codes. In Advances in neural information processing systems, 1410-1418.

[5] Murphy, B., Talukdar, P., & Mitchell, T. M. (2009). Selecting corpus-semantic models for neurolinguistic decoding. In Advances in neural information processing systems, 1410-1418.

[6] Pipanmekaporn, P., Ludmilla, T., Guigue, V., Murphy, B., Talukdar, P., & Artières, T. (2015). Designing Semantic Feature Spaces for Brain Reading. In Proceeding of 23rd European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium.

[7] Bergsma, S. & Goebel, R. (2011). Using Visual Information to Predict Lexical Preference. In Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, 399-405.

[8] Cai, H., Huang, Z., Zhu, X., Zhang, Q., & Li, X. (2015). Multi-Output Regression with Tag Correlation Analysis for Effective Image Tagging. In Proceeding of International Conference on Database Systems for Advanced Applications, 31-46.

[9] Chua, T.S., Jinhui, T., Rihang, H., Haoiie, L., Zhipping, L., & Zheng, Y.T. (2009). NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In Proceeding of ACM International Conference on Image and Video Retrieval, Greece, 1-9.

[10] Ahmad, J., Mehrdad, A., & Khadivi, S. (2010). WordNet Based Features for Predicting Brain Activity associated with meaning of Nouns. In Proceeding of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics. Association for Computational Linguistics, 18-26.

[11] Shapiro, L.G. & Stockman, G.C. (2003). Computer Vision. Prentice Hall.

[12] Huang, S.K., Mitra, M., Zhu, W.J., & Zabib, R. (1997). Image indexing using Color Correlogram. In Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition, IEEE, 762-768.

[13] Park, D.K., Jeon, Y.S., & Won, C.S. (2000). Efficient Use of Local Edge Histogram Descriptor. In Proceedings of the 2000 ACM workshops on Multimedia, 51-54.

[14] Manjunath, B.S. & Ma, W.Y. (1996). Texture Features for Browsing and Retreival of Image Data. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), 837-842.

[15] Josef, S. & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In Proceedings of Ninth IEEE International Conference on Computer Vision, 1470-1477.

[16] Norman, K.A., Polyn, S.M., Detre, G.J., & Haxby, J.V. (2016). Beyond Mind Reading: Multi-Voxel Pattern Analysis of fMRI Data. In Trends in Cognitive Sciences, 10(9), 424-430.

[17] Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., & Rao, A.R. (2009). Prediction and Interpretation of Distributed Neural Activity with Sparse Models. In NeuroImage, 44(1), 112-122.

[18] Daly, J.J. & Huggine, J.E. (2015). Brain-Computer Interface: Current and Emerging Rehabilitation Applications. In Archives of Physical Medical and Rehabilitation, 96(3), 1-7.